CLAIMS

What is claimed is:

1. A computer implemented system that facilitates maximizing probabilities comprising:

a data input component that provides one or more types of data for analysis; and

an analysis component that analyzes at least a subset of one or more types of data to compute maximized probabilities by employing at least one of: an Exponential prior, a LaPlacian prior, or a non-Gaussian distribution and an iterative scaling function.

2. The system of claim 1, the iterative scaling function comprises generalized iterative scaling.

3. The system of claim 1, the iterative scaling function comprises improved iterative scaling.

4. The system of claim 1, the iterative scaling function comprises sequential generalized iterative scaling.

5. The system of claim 1 employing a plurality of Exponential priors, the plurality of Exponential priors corresponding to a plurality of different features, respectively.

6. The system of claim 5, wherein the Exponential prior employed depends on counts of the features.

7. The system of claim of 5, wherein the Exponential prior employed depends in part upon a usefulness of a feature.

8. The system of claim 5, wherein the counts are based in part upon a Good-Turing estimate.

9. The system of claim 1, the analysis component comprising:

a maximization component that provides instructions for computing a maximum value;

a model component operatively coupled to the maximization component that receives data from at least the maximization component and at least an Exponential prior component; and

a probability processing component that employs information collected by the model component to compute one or more values.

10. A computer implemented method that facilitates maximizing probability values comprising:

employing a maximum entropy model using at least one of a plurality of Exponential priors to maximize probability values;

employing an update function for the maximum entropy model, the update function comprising an *observed _ count – discount* term; and

bounding a parameter value.

11. The method of claim 10, bounding the parameter value at 0.

12. The method of claim 10, the plurality of Exponential priors corresponding to a plurality of different features, respectively.

13. The method of claim 10, wherein the Exponential prior employed depends on counts of the features.

14. The method of claim of 10, wherein the Exponential prior employed depends in part upon a usefulness of a feature.

15. The method of claim 13, the counts are based in part upon a Good-Turing estimate.

16. The method of claim 11, the update function comprising:

$$\lambda \leq \max\left(0, \lambda + \frac{1}{n}\ln\left(\frac{observed\_count - discount}{expected\_count}\right)\right)$$

where λ is a parameter and n is a normalizing value.

17. The method of claim 16, n is equal to 1.

18. The method of claim 16, n is equal to $f^{\#}$ which is a maximum sum of features.

19. The method of claim 11, the update function comprises solving for :

$$observed[i] = \sum_j \sum_y P_\Lambda(y \mid x_j)\exp(\delta_i f^{\#}(y, x_j)) + discount$$

20. A computer implemented method that maximizes probability values to facilitate training a machine learning system comprising:

receiving a data set;

determining an Exponential distribution as a prior;

defining one or more parameters; and

training a model based at least in part upon a subset of the data set, the Exponential prior and the one or more parameters.

21. The method of claim 20, determining an Exponential prior comprises:

providing a relatively large data set;

training a model using the large data set and the Gaussian prior;

graphing a distribution of parameter values that have at least 30 training instances; and

determining the Exponential prior by examining the distribution of parameters.

22. The method of claim 20, the Exponential prior being determined based at least in part upon a particular feature of interest.

23. The method of claim 22, the feature is an IP address.

24. A data packet adapted to be transmitted between two or more computer processes facilitating providing suggestions to an online user, the data packet comprising:

information associated with employing a maximum entropy model using at least one of a plurality of Exponential priors to maximize probability values; employing an update function for the maximum entropy model, the update function comprising an *observed _count – discount* term; and bounding a parameter value.

25. A computer-readable medium having stored thereon the following computer executable components:

a data input component that provides one or more types of data for analysis; and

an analysis component that analyzes at least a subset of one or more types of data to compute maximized probabilities at least in part by employing at least one Exponential prior, a LaPlacian prior, or a non-Gaussian distribution and an iterative scaling function.